

METODOLOGIA DO TRABALHO TERMINOLÓGICO: O CASO DA TERMINOLOGIA DO CINEMA EM PORTUGUÊS EUROPEU

Mafalda Antunes (Ait - Associação de informação Terminológica, Lisboa, Portugal)

RESUMEN:

EN ESTE ARTÍCULO SE PRESENTA EL CONJUNTO DE PROCEDIMIENTOS ADOPTADOS EN EL PRECURSO DEL TRABAJO DE DELIMITACIÓN Y ESTABLECIMIENTO DEL VOCABULARIO DEL CINE EN PORTUGUÉS EUROPEO (PE), SIGUIENDO LOS PRESUPUESTOS METODOLÓGICOS PRESENTADOS POR M. TERESA CABRÉ EN LA TEORÍA COMUNICATIVA DA TERMINOLOGÍA (TCT).

LA METODOLOGÍA ADOPTADA PARA IDENTIFICAR LA REFERIDA TERMINOLOGÍA SE ENCUENTRA ESTRUCTURADA EN LAS SIGUIENTES ETAPAS: LAS TAREAS PRELIMINARES; LA CONSTITUCIÓN DEL CORPUS; LA EXTRACCIÓN, ORGANIZACIÓN Y REVISIÓN DE LOS DATOS; Y, AÚN, EL ANÁLISIS Y DESCRIPCIÓN DE LOS DATOS. EN EL PRESENTE ARTÍCULO, SE INTENTA DESCRIBIR CADA UNA DE ESTAS FASES PORMENORIZADAMENTE CON EL OBJETIVO DE PRESENTAR UN TIPO DE TRABAJO QUE SE PONDRÁ APLICAR EN EL ESTABLECIMIENTO DE CUALQUIERA OTRA TERMINOLOGÍA DE OTRA ÁREA ESPECIALIZADA, Y TAMBIÉN EN EL ESTUDIO DEL VOCABULARIO CORRESPONDIENTE.

PALABRAS-CLAVE: TERMINOLOGÍA; METODOLOGÍA; CINE; CORPUS

RESUMO:

NESTE ARTIGO APRESENTA-SE O CONJUNTO DE PROCEDIMENTOS ADOPTADOS NO DECORRER DO TRABALHO DE DELIMITAÇÃO E ESTABELECIMENTO DO VOCABULÁRIO DO CINEMA EM PORTUGUÊS EUROPEU (PE), SEGUINDO OS PRESSUPOSTOS METODOLÓGICOS PROPOSTOS POR M. TERESA CABRÉ NA TEORIA COMUNICATIVA DA TERMINOLOGIA (TCT).

A METODOLOGIA ADOPTADA PARA IDENTIFICAR A REFERIDA TERMINOLOGIA ENCONTRA-SE ESTRUTURADA NAS SEGUINTE FASES: AS TAREFAS PRELIMINARES; A CONSTITUIÇÃO DO CORPUS; A EXTRACÇÃO, ORGANIZAÇÃO E REVISÃO DOS DADOS; E, AINDA, A ANÁLISE E DESCRIÇÃO E DOS DADOS. NO PRESENTE ARTIGO, PROCURA-SE DESCREVER CADA UMA DESTAS FASES PORMENORIZADAMENTE COM O OBJECTIVO DE APRESENTAR UM TIPO DE TRABALHO QUE SE PODERÁ APLICAR NO ESTABELECIMENTO DE QUALQUER OUTRA TERMINOLOGIA DE OUTRA ÁREA ESPECIALIZADA, BEM COMO NO ESTUDO DO VOCABULÁRIO CORRESPONDENTE.

PALAVRAS-CHAVE: TERMINOLOGIA; METODOLOGIA; CINEMA; CORPUS

1. INTRODUÇÃO

Para se identificar a terminologia de uma determinada área de especialidade considera-se essencial o estabelecimento prévio de um conjunto de procedimentos que sirvam de referência e que guiem essa tarefa. Neste artigo pretende-se apresentar o modo como se estruturou o trabalho de identificação da terminologia do cinema em português europeu¹.

A escolha de um trabalho em terminologia e, em particular, na área do cinema deve-se não só à crescente evolução que se verifica na actualidade nos domínios científicos e técnicos e a consequente necessidade de uma comunicação eficaz nas diversas áreas de especialidade; mas também à evolução da técnica cinematográfica em Portugal a par da ausência de uma terminologia sistematicamente elaborada de acordo com critérios terminológicos pré-estabelecidos e estruturados. Por estas razões considera-se que o português europeu (PE) carece de estudos que contemplem as diferentes terminologias de modo a que possa defender o seu papel em cenários de comunicação científica e técnica, tal como refere Colomer Artigas (2002: 401), acredita-se que: “es evidente que una lengua que no esté suficientemente preparada no podrá competir en el escenario de la comunicación multilingüe. Ello implica disponer de un código lingüístico general actualizado y de lenguajes especializados adaptados a las múltiples funciones comunicativas. Para una lengua sin ese nivel de desarrollo no sólo será difícil entrar en el mercado de las industrias y ser utilizada en el intercambio de conocimientos con otras comunidades lingüísticas, sino que además se presenta el riesgo de ser abandonada a favor de otra con más recursos.”

Considera-se que o objectivo principal deste artigo é indicar o modo como se procedeu à identificação da terminologia do cinema em PE, apresentando detalhadamente o conjunto dos procedimentos metodológicos adoptados. Pensa-se que um trabalho com estas características pode contribuir para:

- aprofundar conhecimentos na área da terminologia;
- adquirir técnica de trabalho sistemático em terminologia;
- demonstrar que o quadro teórico adoptado é válido para o estudo de qualquer linguagem especializada;
- responder à necessidade dos utilizadores deste tipo de trabalho terminológico;
- contribuir para que o modelo metodológico deste estudo venha a servir de base a outros estudos relativos a outras áreas de especialidade.

2. METODOLOGIA

A metodologia adoptada neste trabalho divide-se em três fases principais:

Fase 1: Tarefas preliminares

Fase 2: Constituição do corpus

Fase 3: Extracção, organização e revisão dos dados

De seguida será apresentada cada uma destas fases mais pormenorizadamente.

2.1. TAREFAS PRELIMINARES

No âmbito deste artigo considera-se que são tarefas preliminares aquelas que antecedem o trabalho prático propriamente dito. Destas tarefas fazem parte duas actividades distintas e que se complementam: a fase de delimitação e a fase de preparação do trabalho.

Pensa-se que o nível de qualidade que um trabalho terminológico poderá apresentar se encontra directamente relacionado com uma boa delimitação e execução das tarefas preliminares.

2.1.1. DELIMITAÇÃO DO TRABALHO

2.1.1.1. ELEIÇÃO DO TIPO DE TRABALHO TERMINOLÓGICO A REALIZAR

Estabelecer o tipo de trabalho terminológico a desenvolver, com base na situação da análise do domínio antes de se iniciar o trabalho terminológico propriamente dito, é essencial para se poderem delimitar e organizar as fases seguintes.

Nesta fase, foi determinado o número de línguas com que se iria trabalhar, isto é, se se trataria de um trabalho monolíngue, bilingue ou plurilíngue. Estabeleceu-se que o trabalho a realizar seria monolíngue² e trataria o português europeu, uma vez que se pretende estabelecer a terminologia do cinema nesta língua e contribuir para a sua normalização.

O trabalho terminológico pode ser de dois tipos: pontual ou sistemático. Pontual, se o estudo se aplicar a um só termo ou a um conjunto reduzido de termos de uma mesma área, ou mesmo, a um grupo de termos pertencentes a domínios diferentes³; sistemático, se o estudo se aplica ao conjunto dos termos de uma área ou subárea de especialidade, tendo sido este último tipo o eleito para este trabalho.

2.1.1.2. ELEIÇÃO E ANÁLISE DA ÁREA

Na período imediatamente a seguir à eleição do tipo de trabalho terminológico a realizar, elegeu-se o domínio do cinema como área de especialidade com potencial interesse para estudar. Nesta fase, procurou-se identificar o interesse desta pesquisa e saber até que ponto seria necessário e viável o estudo da terminologia do cinema em português europeu.

Considera-se que, tal como refere Brissos (1959: 5), “O cinema é além de uma técnica e de uma arte – a 7ª – como se lhe chama vulgarmente, uma indústria, um comércio, uma ciência.”. Por esta razão e por ser uma área de âmbito alargado, pensa-se ser importante descrever este vocabulário de especialidade, para resolver eventuais problemas de interacção comunicativa dentro das comunidades profissionais de cada subdomínio, de modo a evitar dificuldades de comunicação dentro da comunidade profissional.

No que se refere ao caso específico do cinema português uma primeira abordagem permitiu verificar que há uma forte influência das unidades importadas nos textos deste domínio e que este facto se deve quer à própria história do cinema e ao modo como este surgiu em Portugal, quer pelo modo como este se manifesta artisticamente. Desde a primeira exibição pública das produções dos irmãos Lumière em Paris, até à ascensão de Hollywood, onde surgem os grandes estúdios e onde se centra a principal indústria cinematográfica, o vocabulário do cinema utilizado em Portugal caracteriza-se pela assimilação de muitas das unidades provenientes quer do francês, quer do inglês.

Nesta fase, teve-se, ainda, em conta a análise da realidade profissional com vista à posterior selecção dos assessores, nomeadamente, âmbitos profissionais implicados, organismos que os representam, centros de possível difusão terminológica, etc. O conhecimento da realidade profissional é

bastante útil para a realização do trabalho terminológico, na medida em que permite ter presentes os âmbitos sobre os quais pedir assessoria no decorrer da pesquisa. Também, uma vez acabado o trabalho, serão os profissionais do domínio os principais utilizadores e difusores da terminologia elaborada.

2.1.1.3. SELECÇÃO DOS ASSESSORES

Após a selecção do domínio do cinema e da análise da sua realidade profissional, estabeleceu-se contacto com profissionais da área, de modo a seleccionar possíveis assessores de trabalho, tendo em conta a selecção de diferentes níveis de especialização e âmbitos de representatividade.

Considera-se que um trabalho terminológico deve ser, preferencialmente, realizado por uma equipa mista, formada por terminólogos e especialistas do domínio. No trabalho em equipa, para além dos especialistas, dever-se-á contar também com colaboradores externos, que podem colaborar pontualmente no decorrer do trabalho ou mesmo na sua fase de revisão.

2.1.1.4. ESTRUTURAÇÃO DA ÁREA E DELIMITAÇÃO CONCEPTUAL

A selecção dos assessores, na fase anterior, permitiu uma aproximação à organização conceptual da área, que posteriormente serviu de base para elaborar a árvore do domínio. Foi também nesta fase que se preparou, tal como Cabré (1999: 143-144) indica, a aquisição de “una de las competencias básicas en terminología: la competencia cognitiva. En la realidad se trata de que el terminólogo no especialista adquiera los conocimientos suficientes para que pueda entrar en el tema”.

Só depois de adquirida uma competência mínima da área e de se ter esboçado, ainda que preliminarmente, a árvore do domínio se tornou possível a definição do trabalho a realizar. Os primeiros contactos efectuados com textos do domínio e com especialistas serviram para delimitar, nesta etapa, três grandes fases da técnica cinematográfica: a pré-produção; a produção e a pós-produção.

2.1.1.5 Fixação dos objectivos e delimitação do plano de trabalho

Nesta fase foram determinados o número aproximado⁴ de termos a conter nesta terminologia e as informações de cada ficha terminológica, nomeadamente: a indicação da categoria gramatical, as línguas dos equivalentes, a definição, os contextos, a presença de ilustrações, etc. Determinou-se, ainda, os campos da ficha que seriam de preenchimento obrigatório e aqueles que apenas se preencheriam no caso de se considerar necessário e/ou verificar ocorrência de informação relevante.

Elaborou-se, também, um esboço dos recursos de que potencialmente se necessitaria ao longo do trabalho como: base de dados, leitor óptico, programa de concordâncias, processador de texto, etc.

Tendo em conta os resultados da análise da área, foram estabelecidos os objectivos de pesquisa úteis e tanto quanto possível realistas para cada fase do trabalho, não esquecendo os prazos delimitados, mas também as características do domínio e dos seus destinatários. Apresentam-se, assim, algumas das variáveis deste trabalho⁵, definidas até este período:

Domínio de especialidade	Cinema.
Tema	O vocabulário do cinema em português europeu.
Perspectiva teórica	Teoria Comunicativa da Terminologia.
Tipo de trabalho	Monolingue e sistemático.
Destinatários	Estudantes de cinema, técnicos de cinema, tradutores, jornalistas, etc.
Objectivos	Determinar a terminologia do cinema, apresentar esta terminologia com propostas de normalização, fornecer informação conceptual sobre os termos deste domínio, apresentar e classificar as principais estruturas morfossintáticas presentes neste vocabulário.
Finalidades	Facilitar a redacção técnica, a tradução de textos especializados, corrigir usos não adequados ou equivalentes com baixo índice de fiabilidade, resolver dúvidas sobre a forma e o conteúdo de termos especializados, etc.

Figura 1 – Objectivos e delimitação do trabalho

2.1.2. PREPARAÇÃO DO TRABALHO

2.1.2.1. PREPARAÇÃO DA FUNDAMENTAÇÃO TEÓRICA A SEGUIR NO TRABALHO TERMINOLÓGICO

Os trabalhos de pesquisa têm, normalmente, uma fundamentação teórica baseada em um ou vários modelos teóricos de acordo com o tipo de trabalho que se pretende desenvolver e com as convicções de base de quem o coordena ou realiza. No caso do trabalho terminológico a desenvolver, se

se tratar de um trabalho de equipa, é muito importante que todas as pessoas envolvidas neste processo respeitem esses modelos e os princípios metodológicos adoptados, de modo a haver homogeneidade durante todas as fases.

No caso específico deste trabalho, o contacto prévio⁶ com a Teoria Comunicativa da Terminologia permitiu, à partida, optar pela utilização dos pressupostos defendidos na mesma, por se considerarem os mais adequados e eficazes.

2.1.2.2. ORGANIZAÇÃO CONCEPTUAL DA ÁREA E DEFINIÇÃO DA ÁRVORE DO DOMÍNIO

Após um contacto mais próximo com a documentação disponível da área, procedeu-se à estruturação da árvore do domínio. Esta estruturação procura reflectir a realidade da área com a especificação dos domínios e subdomínios que lhe são próprios e foi feita com a colaboração dos assessores. Tal como considera Cabré (1999: 101), pensa-se que:

Los conceptos (...) no se encuentran aislados dentro de las estructuras del conocimiento, sino que forman parte de conjuntos organizados llamados campos de conocimiento o disciplinas. Y los criterios de organización de los conceptos que integran un mismo campo, así como su pertenencia a un determinado conjunto, están condicionados por la forma en que los objetos de la realidad son comprendidos por los sujetos. Dentro de cada campo, los conceptos se vinculan entre sí sobre la base de dos grandes tipos de relaciones: las lógicas, basadas en la semejanza, y las ontológicas, basadas en la contigüidad o contacto en el espacio y el tiempo.

A árvore do domínio⁷ serve de guia conceptual a toda a pesquisa e concretamente para:

- delimitar a área temática do trabalho;
- fazer um plano sistemático de extracção;
- controlar a pertinência dos termos;
- classificar e ordenar as fichas terminológicas;
- definir as unidades terminológicas da área de maneira lógica e sistemática.

A delimitação da árvore do domínio permite, ainda, distinguir os termos de tronco comum e os de áreas conexas. De facto, a área do cinema é uma área muito próxima de outras áreas de especialidade como a televisão, o teatro, a fotografia, etc. e, por esta razão, poderá haver um uso comum de unidades das áreas envolvidas. O fenómeno de transferência de unidades terminológicas entre áreas de especialidade próximas é bastante comum.

Determinar quais as áreas mais próximas e que podem exercer esta influência na transferência das unidades terminológicas de cada um destes domínios é uma tarefa importante nesta fase, pois nas fases seguintes ter-se-ão estes dados em conta na inserção de textos no corpus. Na figura seguinte⁸ é possível observar as áreas conexas do cinema:



Figura 2 – Áreas conexas à área do cinema.

Nas ontologías, tal como considera Cabré (no prelo b) “cada término está asociado a un concepto ubicado en una estructura de conocimiento que lo vincula a otros conceptos del mismo campo a través de relaciones diversas (hiperonimia, hiponimia, meronimia, holonimia, causalidad, localización, etc.). De

esta forma el contenido de un término representa el conjunto de las relaciones que el concepto asociado a este término establece con el resto de los conceptos del ámbito especializado.”

Apresenta-se, em seguida, uma proposta de representação do sistema conceptual do cinema⁹, que resulta de uma leitura prévia de parte dos documentos do corpus e das consultas aos assessores:

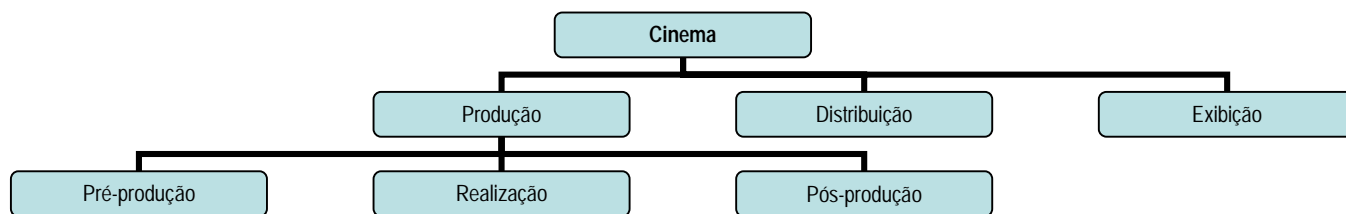


Figura 3 – Sistema conceptual do cinema

Dentro de cada um destes subdomínios principais foi, ainda, possível estabelecer uma tipologia das unidades registadas, que se distribuem do seguinte modo:

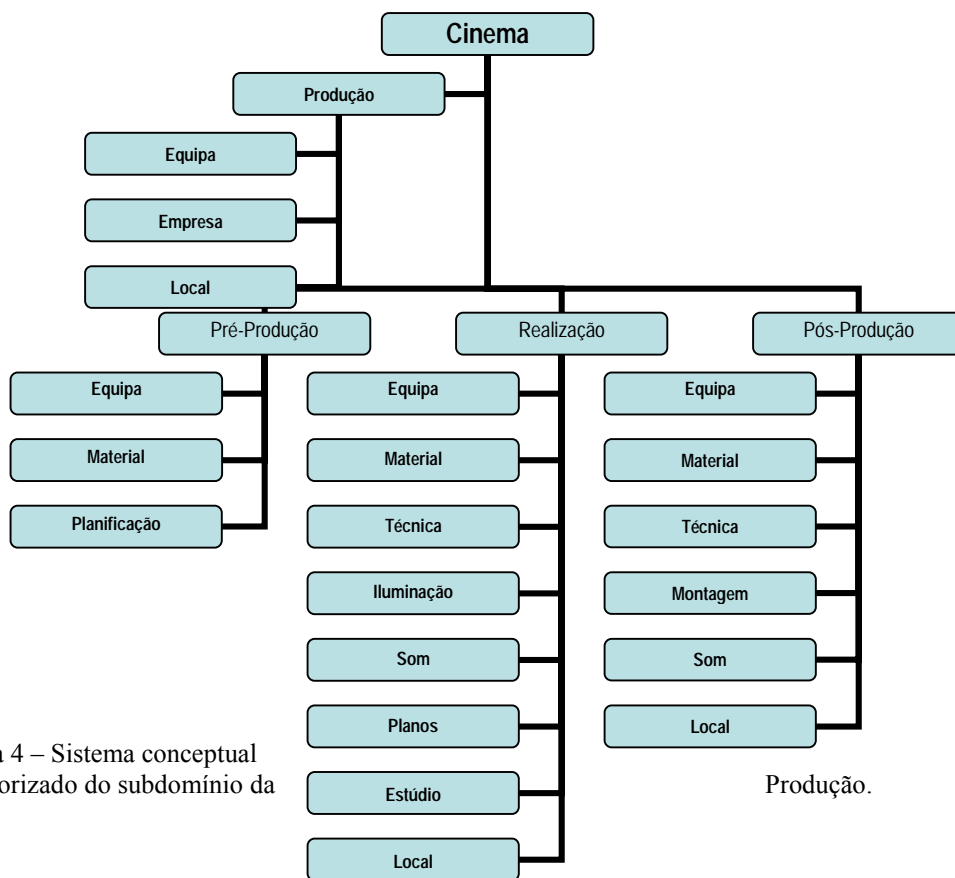


Figura 4 – Sistema conceptual pormenorizado do subdomínio da

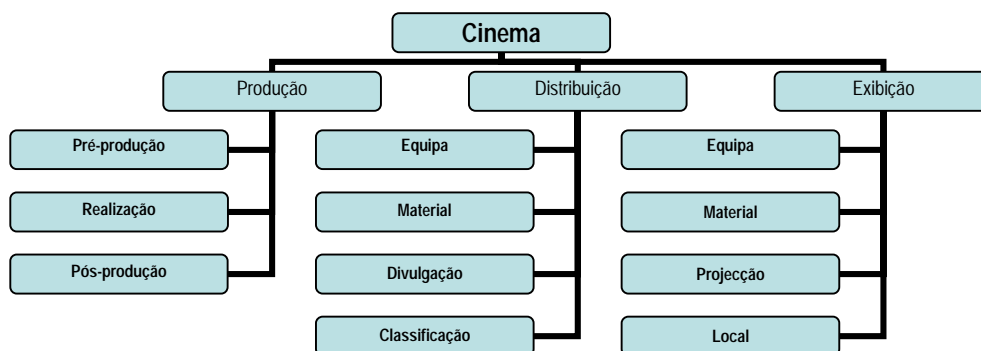


Figura 5 – Sistema conceptual pormenorizado dos subdomínios da Distribuição e Exibição.

De seguida, descreve-se e apresenta-se alguns exemplos de cada um dos subdomínios mais pormenorizadamente, de modo a justificar e estruturação proposta.

I. Produção

A produção cinematográfica consiste na actividade de reunir todos os meios necessários para executar um filme. A equipa de produção, mais especificamente o produtor, é responsável pela organização de todas as fases que consistem na pré-produção, na realização e na pós-produção. As fases da produção compreendem, portanto, as actividades relacionadas quer com o contrato e o estudo do argumento, quer com a preparação do estúdio e dos exteriores, quer com os contratos de trabalho com a equipa técnica e mesmo com a ocupação de laboratórios e estúdios de som.

São exemplos de unidades pertencentes ao subdomínio da produção, unidades como:

Produção – equipa: *chefe de produção, contabilista, produtor executivo;*

Produção – local: *companhia produtora, produtora;*

Produção – empresa: *American International Pictures, Columbia Pictures, Paramount Pictures;*

Produção – outros: *co-produção, orçamento, super-produção.*

– Pré-produção

A fase de pré-produção é a fase inicial de todo o processo de concretização de um filme, começa com a luz verde para a produção do filme e termina no início das filmagens, isto é, na fase da realização. Nesta fase, escreve-se o guião, realiza-se a planificação técnica, escolhe-se o elenco, concebem-se os cenários e prepara-se a organização das filmagens.

São exemplos de unidades pertencentes ao subdomínio da pré-produção, unidades como:

Pré-produção – equipa: *adaptador, autor, escritor, guionista;*

Pré-produção – planificação: *argumento, enredo, guião, story-line;*

Pré-produção – outros: *casting, direitos de adaptação, pesquisa.*

– Realização

A fase da realização consiste no período de direcção das filmagens que integram a obra cinematográfica. Na realização, estão incluídos o trabalho de supervisionar o argumento e a planificação, de

São exemplos de unidades pertencentes ao subdomínio da realização, unidades como:

Realização – equipa: *actor, anotadora, assistente de câmara, cabeleireiro, realizador;*

Realização – estúdio: *cenário, Dreamworks, estúdio (1) ¹⁰, Tóbis, set;*

Realização – iluminação: *baby, fill light, fotografia, luz ambiente;*

Realização – local: *décor, exterior, interior, plateau;*

Realização – material: *acetato de celulose, adereço, bandeira, câmara, dolly;*

Realização – planos: *close-up, contraplano, panorâmica, plano americano;*

Realização – som: *perchar, pitch, pista sonora, som digital, sonorizar;*

Realização – técnica: *acelerado, câmara lenta, digital, focalizar, realizar;*

Realização – outros: *cena final, fotogenia, gravação.*

– Pós-produção

A pós-produção consiste no conjunto das operações que se seguem ao período de realização, como a montagem ou edição e a mistura. É nesta fase que os especialistas da montagem, orientados pelos realizadores, fazem os cortes, colagens e utilizam os recursos técnicos de laboratório que proporcionam os efeitos especiais. Após a montagem, inicia-se a fase de sonorização, executa-se a tarefa de inserir no filme diálogos, ruídos e música gravados antes, durante ou após as tomadas de imagem. De seguida, executa-se a sincronização audiovisual (imagem mais som). No caso de filmes em línguas estrangeiras, são adicionadas as legendas ou feitas as dobragens. Por fim, é feita a tiragem das cópias que passarão à distribuição.

São exemplos pertencentes ao subdomínio da pós-produção, unidades como:

Pós-produção – equipa: *assistente de montagem, director de dobragem, técnico de laboratório;*

Pós-produção – material: *ampliador, copiadora, mesa de montagem, negativo;*

Pós-produção – técnica: *colorização, dissolvência, etalonagem, padronização;*

Pós-produção – montagem: *colagem, editar, salto, sincronizar, tempo de exposição;*

Pós-produção – som: *banda sonora, música original, playback, voz off;*

Pós-produção – local: *estúdio de som, sala de montagem;*
Pós-produção – outros: *interítulos, legenda, versão restaurada.*

II. Distribuição

A distribuição é a fase imediatamente anterior à exibição e posterior à finalização ao nível da pós-produção. Nesta fase a empresa de distribuição compra o filme ao produtor por um determinado período e assegura a sua promoção e gestão. São tiradas as cópias do filme, armazenadas e conservadas para que sejam alugadas a um exibidor.

São exemplos de unidades pertencentes ao subdomínio da distribuição:

Distribuição – classificação: *cinema independente, clássico, longa-metragem, trilogia;*

Distribuição – divulgação: *blockbuster, distribuidor, circuito comercial;*

Distribuição – local: *arquivo, arquivo fílmico, filmoteca;*

Distribuição – outros: *censura, importador, studio system.*

III. Exibição

Nesta fase, o filme chega ao público através do exibidor, que assegura a sua difusão e projecção nas salas de cinema.

São exemplos de unidades pertencentes ao subdomínio da exibição:

Exibição – equipa: *apresentador, crítico de cinema, projeccionista;*

Exibição – local: *bilheteira, drive-in, plateia, sala de projecção;*

Exibição – material: *bilhete, ecrã, projector (1), tela;*

Exibição – projecção: *Dolby Surround, sessão, trailer, velocidade de projecção;*

Exibição – outros: *assistência, espectador, estrear, êxito de bilheteira.*

Foi reunido numa categoria denominada de “Outros” um conjunto de termos que não se inseriam em nenhum dos restantes subdomínios, são exemplos deles:

Outros – corrente artística: *naturalismo, neo-realismo, nouvelle vague, realismo;*

Outros – equipa: *amador, chefe de equipa;*

Outros – tipo de filme: *drama, ficção científica, policial, thriller, western.*

Na classificação dos subdomínios verificaram-se muitos casos problemáticos em que determinados termos poderiam pertencer a mais do que um subdomínio. Nestes casos o critério adoptado, para além da consulta do especialista, foi o de seleccionar o subdomínio mais genérico, para, no caso de se necessitar de especificar, esse trabalho ser mais facilitado. Em cada um dos exemplos apresentados a seguir, optou-se pela classificação através do primeiro sub-domínio, apresentado sublinhado no texto para facilitar a leitura. Exemplos:

equipa de som: realização – equipa ou realização – som;

equipa de iluminação: realização – equipa ou realização – iluminação;

direcção de fotografia: realização – equipa ou realização – iluminação;

pantógrafos: realização – material ou realização – iluminação;

filage: realização – técnica ou realização – planos;

narrador: pós-produção ou pré-produção¹¹.

A classificação dos subdomínios, para além de revelar que os conceitos de um domínio especializado se encontram relacionados entre si de forma sistemática e podem ser descritos ou representados em forma de redes ou estruturas, permite uma abordagem de cariz estatístico dos dados, permitindo, por exemplo, perceber a distribuição do número de unidades recolhidas por subdomínio:

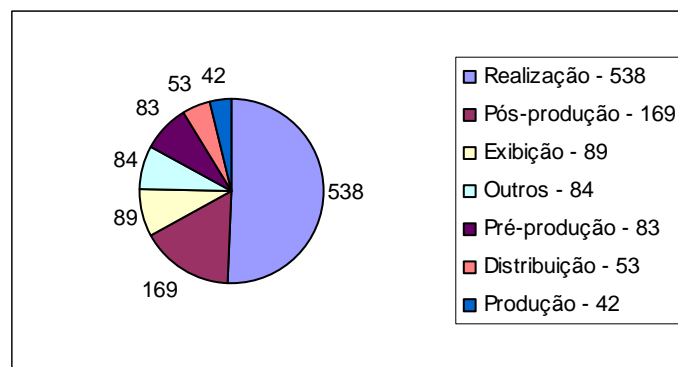


Figura 6 – Distribuição das unidades do cinema recolhidas por subdomínio

Os dados acima apresentados revelam que as fases de realização e pós-produção são, indiscutivelmente, aquelas onde são utilizados mais termos técnicos e, provavelmente, aquelas que necessitam mais de intervenção, nomeadamente no caso da normalização terminológica.

Uma outra informação passível de ser extraída dos dados é, por exemplo, o número de profissionais envolvidos em cada uma das actividades principais do cinema, como revela a figura seguinte:

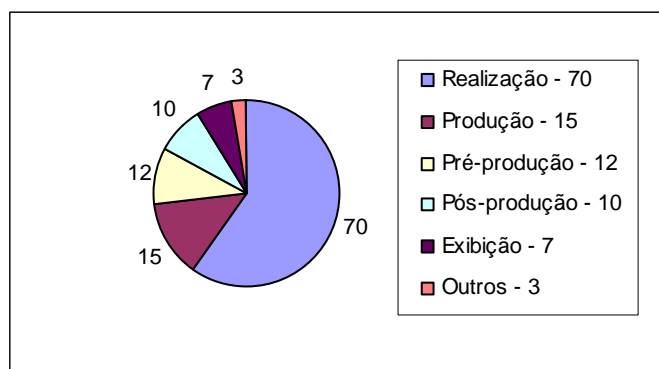


Figura 7 – Distribuição dos nomes de profissão do cinema por subdomínio

2.1.2.3. ELABORAÇÃO DA BASE DE DADOS

Nesta etapa da preparação do trabalho foi elaborada a base de dados na qual foram registados todos os dados pré-definidos relativos a cada unidade terminológica do cinema. Esta base de dados foi criada em Access 2002¹² e os seus campos foram definidos tendo por base todos os dados que se consideram relevantes para uma futura publicação de um dicionário técnico do domínio.

2.1.2.4. SELECÇÃO DAS FONTES A INCLUIR NO CORPUS

Nesta fase, foi feito um inventário de obras em português ou noutras línguas, consideradas úteis para o desenvolvimento deste trabalho. No período da recolha, estabeleceu-se um conjunto de condições às quais deveriam obedecer as obras a ser integradas no corpus, sendo elas:

condição 1: obras de especialidade (incluindo artigos de revistas e jornais especializados);

condição 2: obras editadas em língua portuguesa (incluindo traduções e excluindo o português do Brasil e dos restantes países de língua oficial portuguesa);

condição 3: obras editadas a partir de 1975¹³.

Considera-se que esta é uma das mais importantes fases no desenvolvimento do trabalho terminológico, uma vez que da selecção rigorosa e baseada em critérios rígidos dos textos que constituirão o corpus depende a qualidade dos dados.

É importante, nesta fase, dar uma particular atenção à variedade dos textos, tal como considera Bacelar do Nascimento (2003: 171):

De facto, em corpora especializados, o estabelecimento de critérios externos de selecção dos textos deve supor a inclusão de textos representativos de variados graus de tecnicidade (...). Deve ser sempre assegurado o aval de especialistas à qualidade dos autores de todos estes textos.

A questão da inclusão, ou não, de traduções num corpus especializado é também uma questão importante a ter em conta. Autores como Costa (1993) consideram que não se devem incluir traduções nos corpora; no entanto, no domínio especializado do cinema existem poucos textos originalmente escritos em língua portuguesa. Perante este cenário colocaram-se duas opções: ou não se incluíam textos traduzidos, o que reduziria a extensão do corpus; ou se faria uma selecção rigorosa dos textos traduzidos disponíveis. No âmbito deste trabalho, considerou-se a segunda opção a mais viável, tendo em conta que se seleccionaram traduções de textos com elevado grau de especialização, como é o caso do manual “A realização cinematográfica”, usado por alunos da escola de cinema, tratando-se, portanto, de um livro que é bastante utilizado e que inclui a terminologia técnica veiculada no âmbito do ensino deste domínio em Portugal. A opção pela utilização de traduções teve, no entanto, em conta a fidedignidade do tradutor e da editora.

Na fase de constituição do corpus foram seleccionados, também, textos disponíveis em linha. O peso destes textos no corpus é bastante significativo, uma vez que tal como Wooldridge (2004: 209)¹⁴ considera: “(...) the Web as database has also the enormous advantage over other language corpora of behaving like natural language: it is dynamic, unceasingly renewing itself and thus offering snapshots of the present state of the language, with its proportion of new, established and aging usage.”

Tendo em conta a evolução técnica do domínio em causa, optou-se pela selecção de textos disponíveis em linha, em especial, textos que se referem à técnica digital, pois, dada a datação das fontes não-digitais, não se encontrava disponível informação textual relativa a esta técnica mais recente.

Wooldridge (2004:209), refere, ainda, um factor tido em conta:

Toutefois, comme dans la linguistique de corpus traditionnelle, la valeur des résultats d'une interrogation du Web dépend toujours d'une interprétation intelligente. Il n'est pas douteux que nombre d'aspects de l'usage linguistique ne soient convenablement observables que par les méthodes de la recherche spécialisée. Un des grands avantages du Web-corpus par rapport aux corpus élaborés réside dans la rapidité avec laquelle on peut dépister certains types de phénomènes linguistiques. L'autre avantage majeur du Web réside dans sa nature dynamique. Tout comme la langue elle-même, il se renouvelle sans cesse et offre ainsi des instantanés de l'état actuel de la langue, avec sa dose d'usages nouveaux, établis ou vieillissants.

Outra decisão tomada nesta fase foi a da não inclusão de fontes orais no corpus. Embora se concorde com a afirmação de que “a inclusão de dados do oral num corpus é sempre desejável, mesmo nas linguagens de especialidade, porque sem este registo a descrição linguística carece de completude”¹⁵, neste corpus não foram inseridos registos do oral pois, por um lado, não se obteve autorização para este efeito da parte de nenhum dos assessores e, por outro lado, verificou-se, através da assistência às filmagens, que os dados do oral em cinema apresentam uma grande variedade denominativa. Leech (1991: 11) *apud* Andrade e Correia (2004) refere que “the collection of spoken discourse on the same scale as written text will remain a dream of the future”. De facto esta é também uma das tarefas que não se exclui no desenvolvimento futuro deste trabalho.

A posterior codificação e/ou etiquetagem do corpus, isto é, a inserção de informações (morfológicas, sintácticas, semânticas, discursivas, etc.) relativas a cada unidade do texto foi outros dos factores pensados durante este período. Mesmo tendo em conta que a etiquetagem de corpora pode trazer as mais variadas vantagens na rapidez de exploração dos mesmos, no âmbito deste trabalho, optou-se por não etiquetar o corpus, uma vez que os principais objectivos deste trabalho apenas tinham que ver com a identificação da terminologia e a etiquetagem implicaria um tempo muito superior de aplicação e revisão. No entanto esta hipótese não foi excluída como um dos trabalhos a desenvolver no futuro como consequência deste.

2.2. CONSTITUIÇÃO DO CORPUS

Na fase de elaboração do corpus teve-se sempre em conta que, tal como considera Bacelar do Nascimento (2003: 169), “A qualidade de um projecto terminológico ou terminográfico está directamente relacionada com a qualidade do corpus especializado em que se baseia”.

Nesta fase o material documental recolhido na análise prévia da área foi todo revisto e reorganizado de modo a que esta selecção fosse representativa dos diferentes campos da árvore do domínio. Do material disponível para extracção de unidades terminológicas foi dada prioridade ao mais actual, àquele

que se encontrava escrito em português europeu originalmente e àquele que se considerou ter mais valor técnico e linguístico.

As obras que fazem parte do corpus de extracção, que serão consultadas na fase de preenchimento e revisão do trabalho, foram codificadas, isto é, a cada uma foi atribuído um número que vai do [001] ao [183]. A dimensão do corpus de cinema em português europeu é de 871 136 ocorrências, classifica-se como um corpus aberto, pelo que é possível inserir mais textos, numa fase posterior, sempre que se achar pertinente.

Tal como Bacelar do Nascimento (2003: 169) refere, considera-se que: “(...) dificilmente se consegue constituir um corpus especializado de dimensões muito extensas pelo que, na constituição destes corpora – que, em geral, não ultrapassam 1 milhão de palavras –, se torna imprescindível uma grande exigência no que respeita à qualidade e à representatividade dos textos seleccionados. Só assim se poderá assegurar que sejam fiáveis os resultados das análises e a frequência de ocorrência dos fenómenos linguísticos prototípicos da língua em estudo. (...) Ora a capacidade de introspecção de terminólogos e terminógrafos face à língua que descrevem é geralmente muito reduzida; por isso, ainda mais do que noutras áreas, o sucesso dos seus trabalhos depende, em grande medida, do corpus em que se baseiam.”

A questão da dimensão dos corpora tem sido muito discutida, mas diversos autores têm considerado que o importante é, de facto, a qualidade e não implica necessariamente a quantidade do material disponível. Como afirmam Leech, Garside e Atwell (1983: 25) *apud* Andrade e Correia (2004): “In fact the question «How large?» is meaningless unless it is combined with the question of what different types of text are represented in the corpus”. Sinclair (1991: 9) reforça esta ideia afirmando que “The results are only as good as the corpus, and we are at a very primitive stage of understanding the character of corpora and the relation between decisions on the construction of the corpus and information about the language derived from the corpus.”

No que se refere ao tamanho, Sinclair (1991: 18) considera que “the only guidance I would give is that a corpus should be as large as possible, and should keep as growing.” “Once a corpus is in existence, it needs regular maintenance and upgrading.” Sinclair (1991: 20). De facto, esta parece ser uma contribuição muito importante de Sinclair, manter o corpus em aberto, isto é, actualizá-lo e melhorá-lo sempre que possível, a não ser que se trate de um trabalho diacrónico, limitado a um período específico.

Os modos de constituição do corpus, neste trabalho, foram os seguintes:

- recolha de material em formato electrónico;
- conversão de textos através de digitalização óptica;
- conversão por digitação manual.

Tal como considera Marquez Neto (1995: 84) pensa-se que, após a selecção do material disponível na Internet, “a maneira mais rápida e menos falível de conseguir informatizar material linguístico de maneira a obter a forma informática neutra desejada é através de uma boa leitura óptica desse material.” A digitalização é um processo moroso que se encontra condicionado por diversos factores como a qualidade do leitor óptico ou a qualidade tipográfica do material a digitalizar. Embora a digitalização apresente alguns problemas e exija que, posteriormente, se faça uma revisão, continua a ser uma alternativa mais rápida que a digitação manual de textos.

No que se refere à digitação manual de alguns textos, trata-se de um processo extremamente moroso e é considerado como um último recurso. No entanto, importa referir que, por vezes, esta poderá ser uma opção melhor comparativamente à digitalização, pois há casos, como tipos de fontes muito raras, textos em suportes muito degradados, textos escritos em colunas ou com muitas imagens, em que compensa, em termos de tempo, o recurso a este método.

Considera-se que toda esta fase de elaboração efectiva do corpus é, tal como Sinclair afirma *apud* Marquez Neto (1995: 91):

(...) o equilíbrio de um corpus é «a compromise between what you can get and what you [actually get].

2.3 EXTRACÇÃO, ORGANIZAÇÃO E REVISÃO DOS DADOS

2.3.1. EXTRACÇÃO DOS DADOS

A extracção das unidades terminológicas deste domínio que se consideram termos próprios da área de trabalho com as informações contextuais respectivas foi elaborada de dois modos: manualmente e através da utilização de um programa de extracção semi-automática. A extracção manual foi feita directamente das fontes impressas consultadas, a extracção semi-automática só pôde ser concretizada através do material disponível em formato digital e convertido em formato .txt. No entanto, todos os textos utilizados para a constituição do corpus foram convertidos em formato digital, mesmo aqueles que

foram objecto de extracção manual, uma vez que se utilizaram critérios de frequência para determinar a presença de cada unidade terminológica na terminologia do cinema em português europeu.

2.3.1.1. EXTRACÇÃO MANUAL

A extracção manual foi utilizada na primeira fase de preenchimento da base de dados, embora seja mais morosa que a extracção semi-automática. Trata-se de um método que permite aprofundar o conhecimento da área, questionar a pertinência de determinados textos e da selecção de determinadas unidades, proporcionando, consequentemente, o estabelecimento de um contacto mais frequente com os especialistas. Este tipo de recolha permitiu a sistematização de alguns princípios que sempre devem ser adoptados na identificação de cada unidade terminológica candidata a termo, nomeadamente:

- a sua estrutura, o facto de essa unidade possuir uma estrutura que corresponde a uma unidade lexical de origem ou produto da lexicalização de um sintagma;
- a sua especificidade semântica, o facto de essa unidade possuir um significado específico no domínio ao qual se associa;
- a sua pertinência na estrutura conceptual, o facto de essa unidade ser necessária na estrutura conceptual do domínio do qual faz parte.

O principal problema que se encontra na extracção manual, e nos trabalhos terminológicos em geral, é a dificuldade em distinguir as unidades da língua geral e as unidades das linguagens especializadas. Trata-se de um trabalho que depende da formação adquirida inicialmente e de um conhecimento da área que permita fazer esta distinção, assim como, também depende do contacto com os especialistas que têm um papel fundamental neste processo.

As unidades candidatas a termo foram seleccionadas, depois foi confirmada a sua frequência no corpus através de um programa de concordâncias, e no caso de a sua frequência ser igual ou superior a três ocorrências¹⁶ foram registadas numa ficha terminológica e preenchidos os respectivos campos.

2.3.1.2. EXTRACÇÃO SEMI-AUTOMÁTICA

Após o corpus estar na sua totalidade em formato digital recorreu-se ao método de extracção semi-automática através da utilização do programa Concordance. Trata-se de um programa que permite, através de textos electrónicos de qualquer tamanho, fazer listas de palavras, extrair concordâncias, seleccionar palavras específicas dos textos, verificar os seus contextos, contabilizar o número de ocorrências, verificar as percentagens de distribuição, verificar as unidades (até quatro) que ocorrem à esquerda ou à direita dessa unidade, lematizar uma lista de palavras, etc.¹⁷

Este tipo de extracção apresenta-se como uma boa alternativa à extracção manual, pois permite uma maior rapidez em todo o processo de identificação e selecção das unidades, quer porque se tem acesso às frequências, quer porque permite que se observem todos os contextos e se selecione o que se considera mais adequado para o propósito do trabalho. No entanto, há que ter em conta que qualquer trabalho com estas características exige a verificação por parte do terminólogo e, por esta razão, o conhecimento do domínio é a única forma verdadeiramente eficaz de reconhecer as unidades terminológicas num texto especializado.

Como considera L'Homme (2003: 159), a disponibilidade de corpora electrónicos facilita grandemente a tarefa de recolha de termos e contextos de corpora especializados para aplicações práticas ou de investigação. Leva, ainda, ao desenvolvimento de métodos para encontrar dados terminológicos relevantes. Mas afecta, também, a aquisição de dados das maneiras mais profundas. O uso de técnicas de processamento computacional leva a alguns novos problemas terminológicos; por exemplo polissemia, ambiguidade, combinatórias e variantes de termos não podem ser estudados com tanta confiança. Além disso, têm sido desenvolvidos diferentes métodos para obter informação com parâmetros que já não são baseados em meras sequências de caracteres.

Apresenta-se, de seguida, uma imagem do programa Concordance, com a unidade “cinema”¹⁸, a sua frequência no corpus e selecção de um contexto:

No preenchimento de cada ficha terminológica foram tidas em conta algumas regras de modo a sistematizar o seu preenchimento e a facilitar a fase de revisão. Apresenta-se de seguida algumas dessas regras:

- para cada unidade terminológica foi preenchida uma ficha;
- na primeira fase, a unidade é registada tal como se apresenta no contexto de onde se retirou, a grafia da unidade só é alterada se esta apresentar um erro gráfico;
- na fase de revisão, a unidade é apresentada na sua forma lematizada, isto é, nomes e adjectivos no singular e no género que apresentam e verbos no infinitivo;
- se uma unidade no plural apresenta uma frequência no corpus superior a 75%, comparativamente à mesma unidade no singular, é registada no plural; esse facto fica assinalado na sua categoria gramatical, como nos casos das unidades efeitos especiais (n.m.pl.), bastidores (n.m.pl.), legendas (n.f.pl.), pipocas (n.f.pl.);
- cada unidade escreve-se com letra minúscula, excepto se se trata de um nome próprio ou de uma marca;
- no caso de uma mesma forma ocorrer com duas categorias gramaticais diferentes (nome e adjectivo), são preenchidas duas fichas, porque se trata de duas unidades terminológicas;
- os contextos que trazem alguma informação relativa ao significado do termo são considerados mais úteis terminologicamente que os contextos meramente ilustrativos, que não permitem a delimitação da noção. Cada ficha contém um único contexto, pelo que se regista o mais adequado;
- na classificação do tipo de unidade terminológica, sempre que uma unidade sofre mais do que um processo derivacional é indicado o último;
- as variantes ortográficas encontradas são registadas num campo específico;
- sempre que não se verifica a ocorrência de dados relativos a um dos campos de preenchimento não-obrigatórios, esse espaço fica em branco.

A listagem de unidades recolhidas no corpus do cinema, quer a extracção tenha sido manual¹⁹ ou semi-automática, tem por base critérios de frequência, isto é, cada unidade só é registada numa ficha terminológica se regista uma frequência superior a três ocorrências no corpus, excepto nos casos apresentados em 2.3.3.2. Como este corpus não foi lematizado, quando se verificou o número de ocorrências de cada unidade lexical, através do programa Concordance, apenas se contabilizaram as formas do plural e singular (no caso dos nomes e adjectivos) e no infinitivo (no caso dos verbos), isto é, não se contabilizaram as ocorrências de formas flexionadas dos verbos.

De seguida, apresenta-se a ficha terminológica criada na base de dados:

The image shows a screenshot of a database interface for 'Terminologia do Cinema'. It displays a detailed terminological card for the unit 'guião técnico'. The card is organized into several sections: 'Ficha nº' (Card number) and 'Unidade Terminológica' (Terminological unit) at the top; 'Fonte da UT' (Source of the UT) and 'Subdivisão' (Subdivision) below; 'Categoria gramatical' (Grammatical category) and 'Tipo de UT' (Type of UT); 'Campo de preenchimento' (Filling field) and 'Fonte da ocorrência' (Source of occurrence); 'Contexto' (Context) with a descriptive text; 'Variantes' (Variants) and 'Equivalencias' (Equivalences) with fields for 'Sinónimo' (Synonym), 'Sigla' (Abbreviation), 'Variante ortográfica' (Orthographic variant), and 'Equivalencia em inglês', 'Equivalencia em francês', 'Equivalencia em espanhol'; 'Definição' (Definition) with a detailed description; 'Fonte da definição' (Source of definition); 'Notas' (Notes); and 'Data' (Date) with 'Data de revisão' (Revision date). The card is part of a 'Terminologia do Cinema' database, as indicated by the title on the right. The interface is in Portuguese and includes a search bar and a list of cards on the right side.

Figura 9 – Imagem da ficha terminológica de “guião técnico”.

2.3.3. REVISÃO DA BASE DE DADOS

2.3.3.1. REVISÃO DAS UNIDADES TERMINOLÓGICAS

Na fase de revisão da base de dados deu-se uma atenção particular à forma como aparece cada unidade terminológica, pois não se pretende que esta terminologia seja apenas descritiva e baseada na forma como as unidades ocorrem nos textos, mas que tenha um carácter normalizador. Tiveram-se em conta dois aspectos neste período: a lematização e a hifenização.

Por lematização entende-se, de acordo com o Dicionário de termos linguísticos – volume II (1992: 223) *apud* Galisson & Coste, o “reagrupamento sob uma forma gráfica representativa de todas as formas que pode apresentar uma mesma unidade de significação lexicográfica (tradicionalmente, palavra simples ou complexa). Assim, o infinitivo é, geralmente, escolhido para representar as formas do paradigma verbal, enquanto o masculino singular representa o paradigma nominal e o paradigma adjectival. (...)”.

Cada forma registada no campo dedicado à unidade terminológica foi revista, e no caso de ser necessário, alterada para a forma considerada canónica em cada categoria gramatical. Curiosamente, verificou-se que as formas canónicas são, também, aquelas que atingiram maior frequência dentro dos lemas. As únicas excepções foram os casos em que uma determinada unidade, que pode ocorrer também no singular, aparecia no plural com uma frequência superior a 75% dessa ocorrência no singular.

Foi também dada uma particular atenção à questão da hifenização de unidades derivadas por prefixação e de unidades compostas. Verifica-se uma grande hesitação e assistemática no uso do hífen, pelo que o critério adoptado para selecção, numa fase inicial, foi apenas o da frequência. Posteriormente, estabeleceram-se algumas normas de emprego do hífen, tendo em conta as normas ortográficas vigentes.

2.3.3.2. TRATAMENTO DE CASOS PROBLEMÁTICOS PONTUAIS

Durante a fase de revisão do trabalho, surgiram alguns casos pontuais considerados mais problemáticos que tiveram de ser resolvidos. Esses casos tiveram a ver essencialmente com a eliminação e/ou inserção de fichas terminológicas da base de dados. Registram-se casos em que se tiveram de eliminar fichas (mesmo apresentando uma frequência no corpus superior a três ocorrências), ou casos em que se optaram por inserir fichas (mesmo apresentando uma frequência no corpus inferior a três ocorrências).

Procedeu-se à eliminação de fichas quando não se conseguiu determinar o significado da unidade, nem determinar com segurança a sua pertinência no domínio, como, por exemplo, unidades que ocorrem apenas em títulos ou cabeçalhos sem um contexto satisfatório, ou unidades que fazem parte de outras unidades maiores.

Exemplos desses casos são:

- *Carolco Pictures* apresenta 2 ocorrências²⁰, no entanto, não foi registada porque não se encontraram contextos satisfatórios, isto é, em títulos;
- Os sintagmas: *cinema português* (361 ocorrências), *cinema americano* (65 ocorrências), *cinema brasileiro* (6 ocorrências), *cinema espanhol* (6 ocorrências), embora apresentem uma ocorrência significativa não se consideram unidades terminológicas;
- Os sintagmas: *técnico audiovisual* (3 ocorrências), porque está inserido no sintagma *equipamento técnico audiovisual* e não se refere ao nome do técnico, e *luz incidente* (4 ocorrências), porque é parte da unidade *fotómetro de luz incidente* e não, neste caso, um tipo de luz usada em cinema.

No segundo caso, optou-se pela situação inversa, isto é, foram registadas unidades nas fichas terminológicas com frequência inferior a três, estabelecendo as seguintes excepções:

Excepções	Descrição da excepção	Exemplos
Excepção 1	Unidades que são os equivalentes em português de outras que fazem parte do corpus numa língua estrangeira.	<i>montagem invertida</i> (por <i>flash-back</i>); <i>imagem gelada</i> (por <i>freeze</i>); <i>plano afastado</i> (por <i>long shot</i>).
Excepção 2	Unidades que constituem entradas nos glossários seleccionados.	<i>marcha atrás</i> ; <i>chefe de animação</i> ; <i>angulação</i> .
Excepção 3	Unidades indicadas pelos assessores como fazendo parte da terminologia do cinema.	<i>dissolvência</i> ; <i>luz fria</i> ; <i>motor de corda</i> .
Excepção 4	Unidades recolhidas em registo oral, durante as filmagens ²¹ .	<i>assistente de plateau</i> ; <i>assistente de produção</i> ; <i>filage</i> .

Excepção 5	Unidades que são consideradas sinónimos ou antónimos de outras unidades registadas.	<i>supervisor de montagem</i> (sinónimo de <i>montador e editor</i>); <i>filme cómico</i> (sinónimo de <i>comédia</i>); <i>movimento lento</i> (antónimo de <i>movimento acelerado</i>).
------------	---	---

Figura 10 – Quadro das excepções aplicadas na inserção de unidades na base de dados.

2.3. ANÁLISE E APRESENTAÇÃO DE RESULTADOS

Após o preenchimento da base de dados e a resolução dos casos problemáticos procedeu-se à exploração dos dados recolhidos. A criação da base de dados com recurso ao gestor de base de dados relacionais Access permite uma exploração rápida e eficaz dos dados, através da consulta dos registos existentes, permitindo a criação de tabelas de referência, a visualização dos registos, a escolha de campos a contabilizar, elaboração de gráficos, etc.

3. CONCLUSÃO

Neste artigo pretendeu-se apresentar cada um dos passos dados no decorrer da identificação da terminologia do cinema em PE, as opções que foram tomadas tendo em conta os pressupostos metodológicos da TCT e a natureza dos dados em análise.

Procurou-se, no desenvolvimento deste trabalho terminológico, ter sempre presente a sistematicidade e a coerência no tratamento dos dados, tendo-se sempre em conta, tal como nas palavras de Correia (1999: 273), que “Qualquer classificação implica a assunção de critérios para a sua realização; além disso, qualquer classificação é criticável, dado que implica sempre a valorização de determinados aspectos em detrimento de outros. Por esse facto, é fundamental manter a coerência em relação aos critérios assumidos como garantia da qualidade mínima da descrição a efectuar, tal como é necessário explicitar o mais claramente possível quais são esses critérios”.

Apresentou-se neste artigo o fio condutor do trabalho que leva à identificação de uma terminologia própria de uma área de especialidade. Acredita-se que esta metodologia não se aplica só à terminologia do cinema, mas a qualquer trabalho terminológico que apresente como principais objectivos a identificação de uma terminologia.

Pensa-se que cada domínio de especialidade e cada contacto com os respectivos especialistas, poderá implicar uma reestruturação e adaptação ao longo do percurso de um trabalho. A metodologia aqui apresentada não é, portanto, uma sequência rígida, é, antes, uma linha de orientação, implicando, por vezes, a sobreposição de algumas das tarefas.

As questões metodológicas foram tratadas tendo em conta três períodos principais: i) as tarefas preliminares; ii) a constituição do corpus e iii) a extracção, organização e revisão dos dados.

A fase das tarefas preliminares incluiu o período de delimitação do trabalho e o período de preparação desse mesmo trabalho. Considera-se que o rigor e a coerência foram os factores mais importantes neste período. Após eleição do domínio do cinema destaca-se a selecção dos especialistas, dada a importância que a sua colaboração e disponibilidade representaram para a constituição e sucesso deste trabalho, assim como, toda a fase de organização e exploração conceptual da área que permitiu posteriormente a definição da árvore do domínio.

No que se refere à distribuição dos subdomínios identificaram-se três principais: a *Produção*, que inclui a *Pré-produção*, a *Realização* e a *Pós-produção*; a *Distribuição* e a *Exibição*. Verificou-se que a *Realização* é o subdomínio onde, indiscutivelmente, mais termos técnicos são utilizados, 50,85% do total das unidades que compõem esta terminologia, seguindo-se da *Pós-produção* com 15,97% das unidades. Estes dados eram esperados à partida, uma vez que são estes os subdomínios que exigem um maior e mais rigoroso domínio técnico. As restantes unidades distribuem-se equilibradamente pelos outros subdomínios.

Foi na fase de preparação que se iniciou algum do trabalho prático propriamente dito, como o estudo da fundamentação teórica a seguir, as pesquisas bibliográficas, a construção da base de dados e a selecção das fontes a incluir no corpus, aplicando os critérios pré-estabelecidos.

No que se refere à constituição do corpus, foram apresentados os aspectos que se consideraram dever guiar a selecção de textos a incluir num corpus especializado, tendo em conta que não se pode considerar que um texto é especializado pelo simples facto dele tratar um tema científico (sendo esse um factor importante mas não determinante), mostrando a importância de um texto possuir um sistema conceptual preciso, estabelecido pelos especialistas desse tema e, ainda, que seja reconhecido pela comunidade do domínio como próprio dessa área. Verificou-se que todos os textos especializados se caracterizam pela precisão; no entanto, os textos mais especializados apresentam uma maior tendência para a concisão e um maior uso de léxico opaco onde existe menos variação, tornando-se textos com menor inteligibilidade, mas também com uma maior tendência para a sistematicidade, ao contrário da

maior tendência para a variação denominativa que apresentam os textos menos especializados. Considera-se que a variedade de textos seleccionada contribuiu para o enriquecimento desta terminologia e que o peso dos textos recolhidos na Internet contribuiu para a tornar mais actual. Todos os textos que constituem o corpus foram passados para formato digital e convertidos no formato .txt para que se pudesse usar na fase seguinte um programa de concordâncias que facilitou muito a sua exploração.

De seguida, procedeu-se à extracção dos dados, de forma manual e de forma semi-automática²². Pensa-se que a combinação de ambos os métodos de extracção de terminologia foi bastante produtiva na medida em que permitiu adquirir e experimentar ambas as técnicas de exploração de corpora e perceber as vantagens e desvantagens de cada uma delas. Conforme se foram seleccionando as unidades terminológicas do cinema, foram-se registando em cada uma das fichas terminológicas da base de dados construída para esse efeito.

Por fim, foi efectuada a revisão da base de dados dando particular atenção à lematização e à hifenização das unidades terminológicas e foram resolvidos alguns problemas pontuais que surgiram nesta fase de revisão.

NOTAS:

1. O estabelecimento desta terminologia, assim como a sua análise posterior, fazem parte da minha tese de mestrado intitulada “A terminologia do cinema em português europeu: análise das estruturas mais produtivas” e é no âmbito da mesma que se desenvolveu esta proposta de guia para a estruturação metodológica deste tipo de trabalho. Não se inclui no âmbito do presente artigo a análise desta terminologia, embora a mesma tenha sido efectuada como consequência do trabalho que aqui se apresenta.

2. Embora se registem, sempre que possível, equivalentes em inglês, francês e espanhol na base de dados.

3. Este tipo de trabalho é o que frequentemente se faz na AiT – Associação de Informação Terminológica (www.ait.pt).

4. Inicialmente pensou-se que um número aceitável de termos se situaria entre 500 e 1000 termos, no entanto, atingiu-se um número superior através da aplicação do critério de frequência apresentado (superior a três ocorrências).

5. Cf. Cabré (1999: 144-145).

6. Ao nível do trabalho que desenvolvo na Associação de informação Terminológica (www.ait.pt); através da frequência na *IV Escuela Internacional de Verano de Terminología* no Instituto Universitário de Linguística Aplicada (IULA), Universidade Pompeu Fabra, Barcelona em Julho de 2003 (financiado pelo Instituto de Linguística Teórica e Computacional - ILTEC); e ainda, a durante a fase curricular deste Mestrado, mais especificamente no seminário de *Terminologia e Terminografia*, leccionado pela Prof.^a Doutora Maria Teresa Cabré.

7. De acordo com s/a (1990: 17) *Metodologia del treball terminològic*. (Termcat) Barcelona: Generalitat de Catalunya.

8. Esquema presente na página do sítio da Universidade Lusófona (www.ulusofona.pt).

9. Um dos especialistas consultados (José João Silva) considerou que os subdomínios da “Distribuição” e “Exibição” deveriam ser unidos num só sub-domínio; no entanto, neste caso optou-se por ter em conta a opinião dos restantes especialistas, assim como, a minha opinião pessoal, uma vez que se considera haver uma distinção suficientemente clara entre as duas actividades do domínio cinematográfico.

10. O número após a unidade *estúdio* indica que há mais do que uma unidade com esta forma.

11. Neste caso, optou-se pelo momento efectivo em que o seu trabalho é utilizado, embora a sua utilização fosse determinada na fase de pré-produção.

12. Agradeço a Maria Dória o apoio que me deu na criação desta base de dados, assim como, toda a disponibilidade revelada para o esclarecimento e resolução de alguns problemas que surgiram na fase de preenchimento e actualização da mesma.

13. À excepção do texto com a referência [084] com a datação de 1959.
14. Este excerto foi retirado do resumo do artigo, escrito em inglês.
15. *In*: Andrade e Correia (2004).
16. Este número foi definido em função da verificação no corpus do número de ocorrências de um conjunto de unidades do domínio que constituíam entradas dos glossários consultados ou que foram recolhidas oralmente, deste modo, definiram-se três ocorrências como frequência mínima.
17. Para mais informações sobre o programa consultar o seu sítio em <http://www.concordancesoftware.co.uk/>.
18. Esta é a unidade mais frequente do corpus com 4068 ocorrências (no singular) + 69 ocorrências (no plural).
19. Na extracção manual a frequência das unidades foi verificada na fase de revisão tendo-se eliminado as unidades com frequências inferiores às estabelecidas como critério.
20. Trata-se, portanto, de uma unidade registada durante a recolha manual.
21. Embora os registos orais não façam parte do corpus, estas unidades foram inseridas porque se observou a sua utilização frequente durante as filmagens. Trata-se de unidades que apresentam, no mínimo, uma ocorrência no corpus. No caso de se terem recolhido nas filmagens unidades que não ocorrem nenhuma vez no corpus, não se registaram na base de dados; são exemplos dessas unidades: *lilly*, *pantera* e *clapper*.
22. Através do programa *Concordance*[®].

ABSTRACT:

IN THIS PAPER, A SET OF ADOPTED PROCEDURES IN THE DEVELOPMENT OF THE WORK OF DELIMITATION AND ESTABLISHMENT OF THE VOCABULARY OF CINEMA IN EUROPEAN PORTUGUESE IS PRESENTED, BASED ON THE METHODOLOGICAL GUIDELINES GIVEN BY CABRÉ IN HER COMMUNICATIVE THEORY OF TERMINOLOGY (CTT). THE METHODOLOGY ADOPTED TO IDENTIFY THIS TERMINOLOGY IS STRUCTURED IN THE FOLLOWING MAIN STEPS: PRELIMINARY TASKS, CORPUS BUILDING AND DATA EXTRACTION, ORGANIZATION AND REVIEW; AND THE ANALYSIS OF DATA AND THEIR DESCRIPTION. THE MAIN GOAL OF THIS PAPER IS TO DESCRIBE, IN DETAIL, EACH STEP WITH THE AIM OF PRESENTING A KIND OF WORK THAT CAN BE APPLIED TO THE CREATION OF ANY OTHER TERMINOLOGY FROM A DIFFERENT SPECIALIZED FIELD, AS WELL AS, IN THE STUDY OF THE VOCABULARY.

KEYWORDS: TERMINOLOGY; METHODOLOGY; CINEMA; CORPUS

Referências bibliográficas:

Andrade, Ana M. Rebelo de e Margarita Correia. Desenho de um corpus de especialidade: a propósito do projecto TermiNáutica. *Actas RITerm 1988-2002*, CD-ROM com o ISBN 92-9122-023-X. (VIII Simpósio Ibero-Americano de Terminologia. Cartagena de Índias, Colômbia, Outubro de 2002), 2004.

Antunes, Mafalda e Maria Dória. A difusão da terminologia: a Associação de informação Terminológica (AiT). *Terminómetro – A terminologia em Portugal e nos países de língua portuguesa em África*, p. 28-35, 2005.

Bacelar do Nascimento, M. Fernanda. O papel dos *corpora* especializados na criação de bases terminológicas. *Razões e emoção – Miscelânea de estudos em homenagem a Maria Helena Mira Mateus*. Lisboa: Imprensa Nacional – Casa da Moeda, volume II, p. 167-179, 2003.

Barros, Lidia Almeida. *Curso básico de Terminología*. São Paulo: EDUSP, 2004.

Cabré, M. Teresa. *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Editorial Antártica / Empuréis, 1993.

Cabré, M. Teresa. *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos*. Barcelona: IULA-UPF, (Sèrie Monografies, 3), 1999.

Cabré, M. Teresa. Terminologie et linguistique: la théorie des portes. *Terminologies Nouvelles*. Montréal: Rifa, v. 21, p. 10-15, 2000.

Cabré, M. Teresa. Theories of terminology. Their description, prescription and explanation. *Terminology*, 9:2, pp. 163-199, John Benjamins Publishing Company, 2003.

Cabré, M. Teresa. La Terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro. *Debate Terminológico*, v. 1, [http://www.riterm.net/revista/n_1/cabre.pdf], 2005.

Cabré, M. Teresa, J. Morel e C. Tebé. Propuesta metodológica sobre la detección de relaciones conceptuales en los textos a través de una experimentación sobre la relación causa-efecto. *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: IULA-UPF, 2001.

Cabré, M. Teresa e Rosa Estopà. Unidades de conocimiento especializado, caracterización y tipología. Barcelona: IULA-UPF, 2004 (ms.).

Cardero, Ana María. *El neologismo en la cinematografía mexicana*. México: Universidad Nacional Autónoma de México, Escuela Nacional de Estudios Profesionales Acatlan, 1993.

Cardero, Ana María. *Diccionario de términos cinematográficos usados en México*. México: Universidad Nacional Autónoma de México, Escuela Nacional de Estudios Profesionales Acatlan, 1994.

Colomer Artigas, Rosa. Normalización y normativización en la planificación lingüística. *Terminologia, desenvolvimento e identidade nacional*. Correia, Margarita (org.). Lisboa: ILTEC – Colibri, p. 401-412, 2002.

Correia, Margarita. Neologia e Terminologia. *Terminologia: questões teóricas, métodos e projectos*. Mateus e Correia (orgs.). Lisboa: Publicações Europa-América, p. 59-74, 1998.

Costa, M. Rute. *Terminologia da economia monetária*. Dissertação de Mestrado. Lisboa: Universidade Nova de Lisboa – Faculdade de Ciências Sociais e Humanas. (inérita), 1993.

Feliu, Judit. Propuesta de clases conceptuales y de relaciones conceptuales: recopilación y análisis. *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: IULA-UPF, 2001.

Franquesa, Ester. La terminología en el desarrollo y la pervivencia de las lenguas. *Terminologia, desenvolvimento e identidade nacional*, Margarita Correia (org). Lisboa: Edições Colibri, p. 473-480, 2002.

L’Homme Marie-Claude, Ulrich Heid e Juan C. Sager. Terminology during the past decade (1994-2004) – An editorial statement. *Terminology*, 9:2, pp. 151-161, John Benjamins Publishing Company, 2003.

L’Homme Marie-Claude. *La terminologie: principes et techniques*. Montréal: Les Presses de l’Université de Montréal, 2004.

Marquez Neto, Paula. *Combinatórias lexicais no discurso da astronomia – Um estudo em estatística lexical*. Dissertação de Mestrado em Linguística portuguesa Descritiva apresentada à FLUL, Lisboa (inérita), 1995.

Mateus, Maria Helena. Terminologia em Portugal: necessidades em matéria de ordenamento terminológico. *Terminómetro – A terminologia em Portugal e nos países de língua portuguesa em África*, p. 9-12, 2005.

Pavel, Silvia e Diane NOLET. *Manual de Terminologia*. Adaptação para língua portuguesa de Enilde Faulstich. Canadá: Translation bureau / Bureau de la traduction – Direcção de terminologia e normalização, departamento de tradução do governo canadense, 2002.

Pearson, Jennifer. *Terms in context*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 1998.

Sinclair, John. *Corpus, Concordance, Collocation*. Oxford: OUP, 1991.

Termcat. *Metodologia del treball terminològic*. Barcelona: Generalitat de Catalunya. Departament de Cultura, 1990.

Wooldridge, Russon. Le web comme corpus d'usages linguistiques. *Cahiers de lexicologie – Revue internationale de lexicologie et lexicographie*, v. 85, 2004-2, p. 209-225, 2004.

Mafalda Antunes é mestre em Linguística Geral (Lexicologia e Lexicografia), pela Faculdade de Letras da Universidade de Lisboa, com a dissertação “A terminologia do cinema em português europeu – análise das estruturas mais produtivas”. É investigadora no ILTEC – Instituto de Linguística Teórica e Computacional, desde 2002, e desenvolve actividade no âmbito do recenseamento e divulgação da produção terminológica realizada em língua portuguesa na AiT – Associação de Informação Terminológica.

Para além do trabalho de investigação desenvolvido em diversas áreas de especialidade (Economia, Internet, Náutica, Comércio Electrónico, Cinema, etc.), colaborou em vários projectos terminológicos, nomeadamente, *Neologismos da Economia*, *Glossário Pan-Latino do Comércio Electrónico* e *Dicionário de Termos do Comércio Electrónico*. Actualmente, colabora no ONP – Observatório de Neologia do Português Europeu e apresentou o seu projecto de doutoramento nesta área.